Report on Numerical tours

Sebastiano Scardera

1 Tour on Linear programming

1.1 Optimal Transport of Discrete Distributions

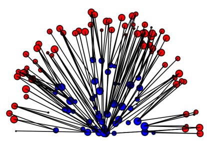
The first part of the notebook is about to solve the Kantorovich problem between two discrete distributions:

$$P^\star \in \text{argmin}_{P \in U(a,b)} \sum_{i,j} P_{i,j} C_{i,j},$$

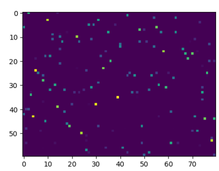
where $C_{i,j}$ is the Euclidean distance between the point x_i and x_j and $U(a,b)$ is defined as follow:

$$U(a,b) := \left\{ P \in \mathbb{R}_+^{n \times m} : \forall i, \sum_j P_{i,j} = a_i, \forall j, \sum_i P_{i,j} = b_j \right\}.$$

We chose a data distributions and putted on it two random discrete measures. We solved it using CVXPY, obtaining the following optimal coupling:



(a) Optimal coupling



(b) Sparsity of the Optimal coupling

One can show that the optimal coupling is sparse; in particular one can check that the optimal coupling has less then n+m-1 non zero entries. The Figure 3b shows the sparsity of the solution.

1.2 Displacement Interpolation

Since the W_2 distance is a geodesic distance, this geodesic path solves the following variational problem

$$\mu_t = \text{argmin}_\mu (1-t) W_2(\alpha, \mu)^2 + t W_2(\beta, \mu)^2,$$

so the map $t \in [0,1] \mapsto \mu_t$ shows the most convenient path from barycenter to barycenter. In our case that $\alpha = \delta_x$ and $\beta = \delta_y$, one has $\mu_t = \delta_{x_t}$ where $x_t = (1-t)x + ty$.

Once the optimal coupling $P^\star$ has been computed, the interpolated distribution is obtained as

$$\mu_t = \sum_{i,j} P^\star_{i,j} \delta_{(1-t)x_i + ty_j}.$$

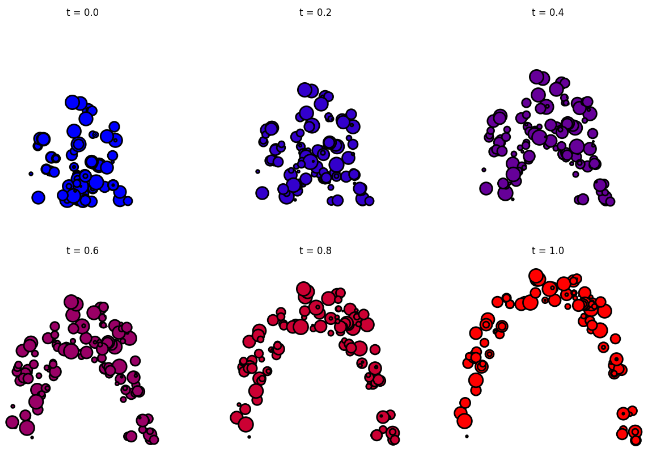
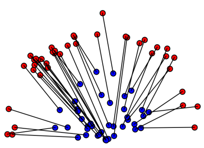


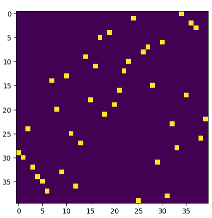
Figure 2: Evolution of μ_t varying t

1.3 Optimal Assignment

When $n = m$ and the weights are uniform, i.e., $a_i = \frac{1}{n}$ and $b_j = \frac{1}{n}$, it can be demonstrated that there exists at least one optimal transport coupling that takes the form of a permutation matrix. This result arises from the property that the extreme points of the polytope $U(1,1)$ are permutation matrices. The following plots shows our experiments:



(a) Optimal coupling



(b) Sparsity of the Optimal coupling

2 Tour on Entropic Regularization of Optimal Transport

2.1 Entropic regularization

We work with two input histograms $a, b \in \Sigma_n$, where we define the simplex in $\mathbb{R}^n$ as:

$$\Sigma_n \equiv \{a \in \mathbb{R}_+^n : \sum_i a_i = 1\}.$$

We examine the following discrete entropically regularized transport problem:

$$W_\epsilon(a,b) \equiv \min_{P \in U(a,b)} \langle C, P \rangle - \epsilon E(P).$$

Here, the polytope of couplings is given by:

$$U(a,b) \equiv \{P \in (\mathbb{R}_+)^{n \times m} : P\mathbf{1}_m = a, P^\top \mathbf{1}_n = b\},$$

where $\mathbf{1}_n \equiv (1, \ldots, 1)^\top \in \mathbb{R}^n$.

For $P \in \mathbb{R}_+^{n \times m}$, its entropy is defined as:

$$E(P) \equiv -\sum_{i,j} P_{i,j}(\log(P_{i,j}) - 1).$$

The regularized transport problem can also be reformulated as a projection:

$$W_\epsilon(a,b) = \epsilon \min_{P \in U(a,b)} \mathrm{KL}(P \| K) \quad \text{where } K_{i,j} := e^{-\frac{C_{i,j}}{\epsilon}},$$

of the Gibbs kernel K according to the Kullback-Leibler divergence.

2.2 Bregman Projection Algorithm and Sinkhorn's Algoirthm

Given two affine constrain set (C_1, C_2), the Bregman Projection algorithm consists in alternating project on this constrain.

The Sinkhorn's Algorithm exploit this idea and the fundamental remark that the optimal coupling P_ϵ has the form:

$$P_\epsilon = \operatorname{diag} u K \operatorname{diag} v$$

where the Gibbs kernel is defined as

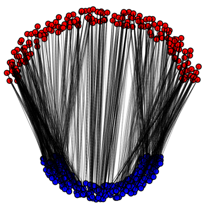
$$K := e^{-\frac{C}{\epsilon}}.$$

Sinkhorn's algorithm alternate between the resolution of these two equations, and reads

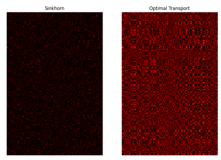
$$u \longleftarrow \frac{a}{Kv} \quad \text{and} \quad v \longleftarrow \frac{b}{K^\top u}.$$

2.3 Transport Between Point Clouds

As in the first numerical tour, We chose a data distributions and putted on it two random discrete measures. The cost matrix is one given by the Euclidean distance. So, the Gibbs kernel is the Gaussian one. We solved the Entropic regularized problem using both Sinkhorn's Algorithm and CVXPY. In Figure 4b we can see the difference between the two obtained couplings. In particular, we can notice how to one from Sinkhorn is much more sparse.



(a) Sinkhorn coupling



(b) Comparison between the two coupling

We can notice also that decreasing ϵ one obtain sparser and sparser solution (Fig. 5

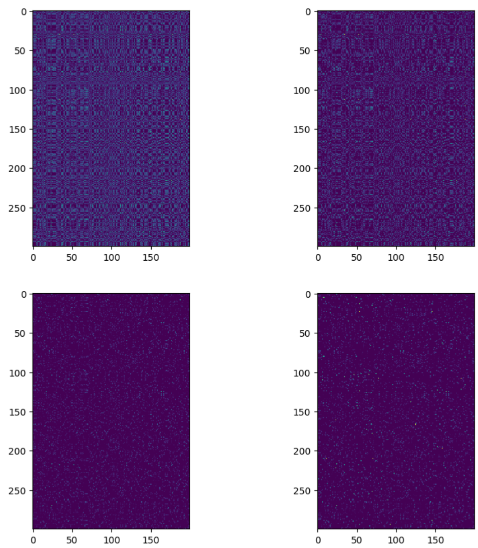


Figure 5: Different solutions obtained decreasing ϵ

2.4 Transport Between Histograms

We now consider a different setup, where the histogram values a, b are not uniform, but the measures are defined on a uniform grid $x_i = y_i = i/n$. They are thus often referred to as "histograms".

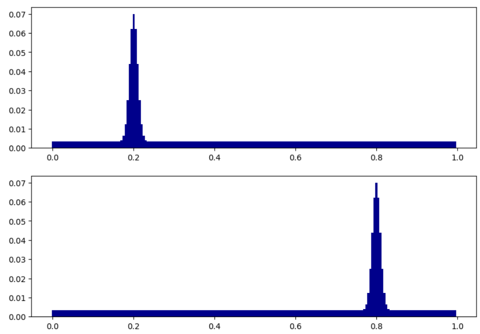


Figure 6

We applied Sinkhorn algorithm to the Entropic regularized problem. The resulting coupling (in log domain) can be seen in Fig. 7a.

(a) (b)

One can compute an approximation of the transport plan between the two measure by computing the so-called barycentric projection map

$$t_i \in [0,1] \longmapsto s_j := \frac{\sum_j P_{i,j} t_j}{\sum_j P_{i,j}} = \frac{[u \odot K(v \odot t)]_j}{a_i}.$$

where $\odot$ and $\div$ are the entry-wise multiplication and division.

This computation can thus be done using only multiplication with the kernel K. The transport map is visible in Fig. 7b.

3 Wasserstein Barycenters

Instead of focusing on computing transport, we now shift our attention to the task of determining the barycenter of R input measures $(a_k)_{k=1}^R$. A barycenter b is defined as the solution to the following optimization problem:

$$\min_b \sum_{k=1}^R \lambda_k W_\gamma(a_k, b),$$

where λ_k are positive weights satisfying $\sum_k \lambda_k = 1$.

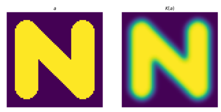
In this specific case, the kernel K associated with the squared Euclidean norm is a convolution with a Gaussian filter:

$$K_{i,j} = e^{-\|i/N - j/N\|^2/\epsilon},$$

where (i, j) are 2-D indexes.

The multiplication against the kernel, i.e., $K(a)$, can now be computed efficiently using fast convolution methods. We use here the fact that the convolution is separable to implement it using only 1-D convolution, which further speeds up computations. In Figure 8b the application of the kernel can be seen.

(a) Input histograms



(b) Application of the kernel

Figure 8

The problem of barycenter computation reduce to optimizing over couplings $(P_k)_{k=1}^R$, and that this can be achieved using an iterative Sinkhorn-like algorithm, since the optimal coupling has the scaling form:

$$P_k = \mathrm{diag}(u_k) K \,\mathrm{diag}(v_k),$$

for some unknown positive weights (u_k, v_k).

The first iteration is the same as Sinkhorn, while the second step of the Bregman projection consists in:

$$\log(b) \equiv \sum_k \lambda_k \log(u_k \odot K(v_k)),$$

After 1600 iterations the barycenter is: